Proceedings of the International Conference of Machine Vision and Machine Learning
Prague, Czech Republic, August 14-15, 2014
Paper No. 67

Fast Superpixel Segmentation Using Morphological Processing

Wanda Benesova, Michal Kottman
Faculty of Informatics and Information Technologies
Slovak University of Technology, Bratislava, Slovakia
benesova@fiit.stuba.sk

Abstract- Superpixels are segments in an image which can serve as basic units in the further image processing. Their purpose is to reduce the redundancy in the image and increase efficiency from the point of view of the next processing task. Many methods for the computation of superpixels were already presented. A drawback of most of these methods is their high computational complexity and hence high computational time consumption. Watershed segmentation seems to be an appropriate fast way for superpixel segmentation, but it is necessary to remove the noise and local extremas that cause unwanted oversegmentation. Second, a sophisticated method for marker image calculation which respects both the remaining natural edges in the image and the regularity of marker placement in still regions of the image still has to be developed. This paper provides a fast method of segmenting an image into superpixels by a morphological approach. First we exploit morphological image reconstruction to eliminate irrelevant spatial local extreme intensities in the image and hence also to remove irrelevant edges. Then we generate markers for morphological watershed segmentation. The result is a fast algorithm comparable to state-of-the-art algorithms that is convenient for real-time or near real-time applications.

Keywords: Superpixels, Segmentation, Morphological processing.

1. Introduction

Superpixels are regions in an image which can be used as basic units (primitives) in the next image processing like segmentation, salience mapping or object detection. Superpixels typically cover the whole image, they are distributed regularly with respect to the nature of the input image, the desirable variation of the size of superpixels is preferably small and the boundary of superpixels has to be corresponding with the natural boundary of objects presented on the image.

Our purpose is to propose a computational efficiency superpixel segmentation method which can be used in near real-time applications for object segmentation and object recognition. The basic concept assumes the segmented superpixels as the basic regions for calculation of correspondent local area features and then those processing in feature space using a classification method.

This paper presents a novel fast method of the dense over-segmentation method using methods of morphological processing especially the morphological reconstruction introduced by Vincent (Vincent, 1993) and morphological watershed segmentation, both controlled by markers.

2. Related Work

Selected published methods of superpixel segmentation are shortly summarized in this section.

In Spatially Coherent Clustering Using Graph Cuts (Zabih and Kolmogorov, 2004), Zabih and Kolmogorov propose a method with the goal to overcome the absence of spatial coherence in segmentation if a clustering in feature space is used. A energy function which consists of a term representing the energy in the spatial space and a term representing the energy in the feature space is to be minimized using graph cuts.

Veksler and Boykov (Veksler et al., 2010) formulate the superpixel partitioning problem in an energy minimization framework, and optimize with graph cuts. Presented energy function explicitly encourages regular superpixels and this method is also suitable for 3D "supervoxel" segmentation. An image is covered with overlapping square patches of fixed size. Hence, each pixel is covered by several patches, and the task is to assign a pixel to one of them.

TurboPixels (Levinshtein et al., 2009) is an iterative algorithm which starts by evolution from seeds placed regularly in the image. The algorithm then iterate until no further evolution is possible, i.e., when the speed at all boundary pixels is close to zero. The iteration loop involve: an evolution of this boundary, estimation of the skeleton of the unassigned region and updating of the the speed of each pixel on the boundary and of unassigned pixels in the boundarys immediate vicinity.

Shi and Malik (Shi and Malik, 1997) propose a graph-theoretic criterion for measuring the goodness of an image partition - the normalized cut. The authors showed that the minimization of this criterion can be formulated as a generalized eigenvalue problem. A computational method based on this idea has been developed and presented by the authors and applied to segmentation of brightness, colour, and texture images.

Felzenszwalb and Huttenlocher (Felzenszwalb and Huttenlocher, 2004) define a predicate for evaluating of two regions of an image whether or not there is evidence for a boundary between two components in a segmentation. This predicate is based on measuring the dissimilarity between elements along the boundary of the two components relative to a measure of the dissimilarity among neighboring elements within each of the two components.

Achanta (Achanta et al., 2010) present two segmentation methods which may be assigned in the category of gradient ascent algorithms. SLIC Simple Linear Iterative Clustering method clusters pixels on their colour according to pixels similarity and proximity in the image plane. The authors use 5-dimensional vector, which consists of the pixel colour vector in CIE L*a*b* colour space and spatial coordinates x, y using Euclidean distances for the measure of similarity.

3. Proposed Method - Morphological Superpixel Segmentation (MSS)

Our goal is to use the superpixels as basic units in the salient object detection and in the object recognition task in our future work. With respect to this goal the desired method is expected to fulfil the following criteria:

- The method should be fast enough to run in near real time application.
- The clusters should have low intra-cluster variation and/or high inter-cluster variation.
- Spatial coherence of clusters in image space.
- The boundary of the segmented superpixel should follow the boundary of regularly sized and regularly distributed rectangular regions as far as no saliency edge are in the neighbourhood.

Local spatial maxima and local minima in image, whose area is small compared with the desired size of the segmented superpixel (irrespective the corresponding intensity value of this local extrema) will be regarded as irrelevant and will have no contribution to the segmentation.

Our approach is based on morphological processing and use of 8-connectivity. This methods are time efficient and guarantee a spatial coherence. The crucial point is a question: which edge is a saliency edge and which edge should be ignored in the process of superpixel boundary? For this decision we have constructed an enhanced image which has been done by removing of all regional-local intensity extrema, hence removing all edges with local significance. The remaining edges are global and will be accepted as saliency edges.

3.1. Image Enhancement Using Morphological Reconstruction

Vincent (Vincent, 1993) has introduced a morphological reconstruction which belongs to geodesic operations. The operation finds all local extrema - peaks. The morphological process of removing of this detected peaks from image f is controlled by a marker image g.

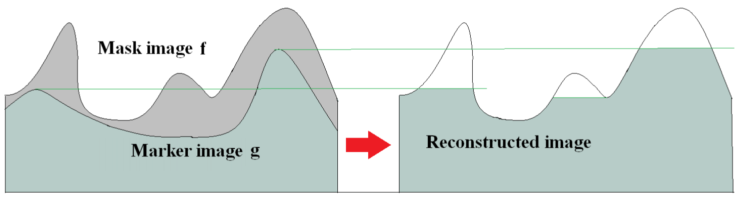

Fig. 1. Morphological reconstruction of a 1D signal.

Given an input image f, every value of the marker image g has to satisfy Equation 1 in each x, y position. The gray-scale reconstruction $p(J)$ of f from g for a pixel p (with thresholds T_k) is given by Equation 2.

$$g(x,y) \leq f(x,y) \tag{1}$$

$$p_I(J)(p) = \max\{k \in [0, N-1] | p \in p_{T_k(I)}(T_k(J))\} \tag{2}$$

The operation of morphological reconstruction is illustrated in Figure 1. The proper marker image strongly depend on the desired application, in our case should marker image allow a removing of local extrema - circular blobs with given maximum diameter. For this purpose we use a marker image generated simply by morphological erosion. The parameter — size of circular structural element S — can be directly derived from the maximum area diameter of local peaks to be removed.

We experimented with multiple values of S and selected $S = 7$. We apply the operation of morphological reconstruction 6 times: $3\times$ for input image R,G,B channels (removing local maximum extrema) and $3\times$ for R,G,B channels of inverted image (removing local minima extrema). The result of applying morphological reconstruction on an image is illustrated in Figure 2.

Fig. 2. Left: Input image. Right: Morphological reconstruction removes local extrema.

3.2. Marker Generation for Watershed Segmentation

For segmentation we use the common technique — marker controlled morphological watershed segmentation originally proposed by (Beucher and Lantuejoul, 1979). The process of generating the watershed markers is of high importance in this case. The markers (seeds) have to respect the regular distribution of superpixels on on the one hand and saliency edges on the other hand. Already presented watershed superpixel segmentations are based only on regular distribution of markers. We use this basic regular superpixel segmentation as reference benchmark in our evaluation.

The previously described morphologically enhanced image (see Section 3.1) is used as an input for the next calculation — difference of Gaussian-blurred image and the input image. The output difference image is then blurred one more time with second Gaussian kernel convolution. The output of this processing is a gradient edge image E used to place the markers seeds as follows:

- each seed is first placed in the local minimum of the rectangular area given by regular grids

- the seeds are growing by morphological flooding (using 8-connectivity) in the image E while this growing area is restricted by the rectangle area of regular grids. The flooding threshold ft is a parameter for balancing between contribution of the edges in image E and contribution of regular grids.

The effect of varying the threshold ft is illustrated in Figure 3.

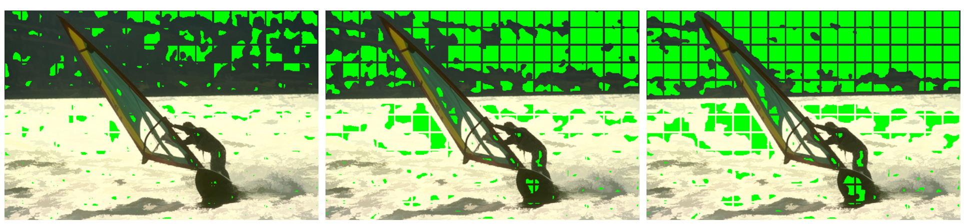

Fig. 3. The effect of choosing threshold ft for marker generation, left to right: 0, 1, 2.

3.3. Marker controlled Watershed segmentation

Meyer proposed an efficient algorithm for marker controlled watershed segmentation based on growing regions in (Meyer, 1992) based on the similarity measure d between a point p and its neighbouring marked region r is colour difference:

$$d = \max\{|p_r - r_r|, |p_g - r_g|, |p_b - r_b|\} \tag{3}$$

This definition was used in the presented evaluation, but we plan to investigate more sophisticated similarity measures in our future work.

The entire sequence of operations of MSS can be seen in Figure 4.

4. Evaluation

In order to evaluate the superpixel segmentation algorithms, we decided to modify the Superpixel Benchmark Toolbox (Neubert and Protzel, 2012), which was used to evaluate multiple superpixel segmentation algorithms: Normalized Cuts, Felzenszwalb-Huttenlocher Segmentation, Edge Augmented Mean Shift, Quickshift, Marker-Controlled Watershed Segmentation (WS), Entropy Rate Superpixel Segmentation, Turbopixel Segmentation, and Simple Linear Iterative Clustering (SLIC).

To verify that our algorithm achieves comparable performance to the benchmark, we also independently evaluated the WS and SLIC algorithms. The results of these algorithms are present in our evaluation and they correspond to the results obtained in (Neubert and Protzel, 2012). All algorithms were evaluated on the

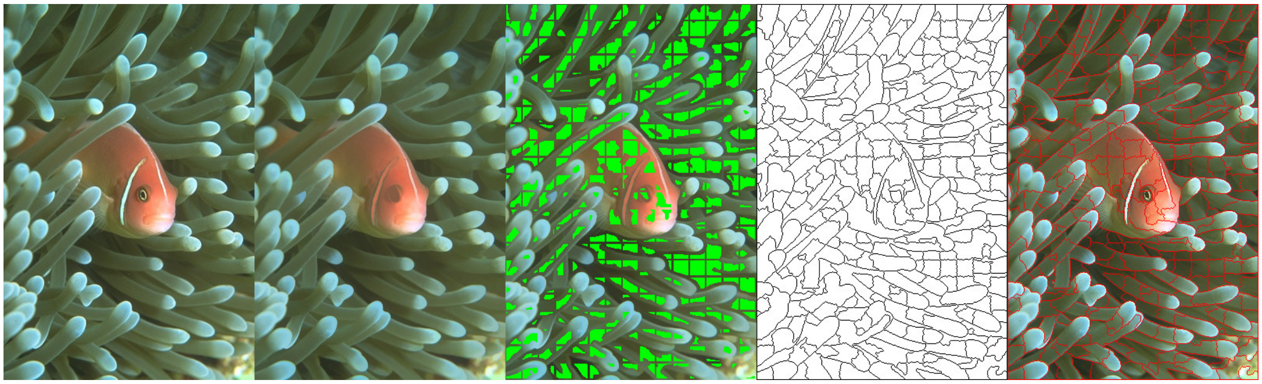

Fig. 4. MSS in steps, left to right: input image; morphological reconstruction; markers for watershed; superpixel boundaries after watershed; fused image.

Berkeley Segmentation Data Set and Benchmarks 500 (BSDS500) dataset (Arbelaez et al., 2011), containing 500 images and hand-drawn contours serving as ground truth.

We used four metrics to evaluate the different superpixel segmentation algorithms: Boundary recall, Undersegmentation error, Mean distance to edge and Intra-cluster variation.

The first two metrics were used verbatim from the Superpixel Benchmark Toolbox in order to achieve comparable results. They require a ground truth segmentation made by humans, available in the dataset. We have proposed the remaining two metrics with the goal to evaluate additional aspects of the segmentation.

All metrics were evaluated for each image (of the dataset of 500 images) processed with a given algorithm to a requested number of superpixels (the resulting number of superpixels can be different than requested). For comparison, we also added a simple algorithm which creates superpixels by means of simple rectangular areas in a rigid grid (BOX).

4.1. Boundary Recall

Boundary recall (BR) is the fraction of hand-segmented edges which lie within a threshold distance k of any superpixel edge (in our experiments, $k = 2$). Since there can be multiple ground truth images for a single input image, they are added together using the OR operation.

The *true positives* (TP) count is the number of pixels in hand-segmented image, for which there is a superpixel boundary pixel in range k. The *false negative* (FN) count is the number of pixels in hand-segmented image for which there is no superpixel boundary pixel in range k. Given these, we can calculate the boundary recall BR as in Equation 4:

$$BR = \frac{TP}{TP + FN} \tag{4}$$

The disadvantage of this metric is that it does not take into account the direction of the edges. Superpixel borders which intersect hand-segmented edges also contribute to the boundary recall. This metric also does not distinguish between superpixel edges which are off by 0, 1 and 2 pixels — they all contribute to the boundary recall equivalently.

4.2. Undersegmentation Error

Undersegmentation error (UE) describes how much area of superpixels crosses the hand-segmented edges. Please refer to original paper (Neubert and Protzel, 2012) for more information on its calculation.

4.3. Mean Distance to Edge

The main purpose of mean distance to edge (MDE) metric is to solve the issues with boundary recall, mainly superpixel–hand-drawn segmentation intersections contributing positively to the recall.

To calculate MDE, we first apply distance transformation to the superpixel segmentation, to get a value for each pixel specifying how far it is from any superpixel edge. Then, we process the hand-segmented image by summing all the distance values for non-zero pixels in the ground truth data. Given a distance image $dist$ and a list of hand-segmented pixels HS, the calculation of MDE can be seen in Equation 5:

$$MDE = \frac{1}{N} \sum_{p \in HS} dist(p) \tag{5}$$

The main motivation behind this metric is to favour such segmentation, where the superpixel boundaries follow the human segmentation more closely. With BR, it is sufficient for a superpixel boundary to lie within k pixels of the hand-segmentation to count as TP. With MDE, it would get a higher mean distance than a boundary which exactly follows the hand-segmentation.

4.4. Intra-cluster variation

This metric describes the quality of segmentation by calculating the mean standard deviation within each superpixel (cluster of pixels). Good segmentation should create homogeneous clusters with smaller differences within each superpixel. In order to compare the intra-cluster variation, we measure the standard deviation of RGB values within each superpixel, and average it over the entire image.

Given a set of superpixels S in an image, where each superpixel s is a set of pixels belonging to it and having a mean value of μ_s, we can calculate IV as in Equation 6:

$$IV = \frac{1}{|S|} \sum_{s \in S} \frac{\sqrt{\sum_{p \in s} (p - \mu_s)^2}}{|s|} \tag{6}$$

5. Results

5.1. Evaluation Using Metrics

The results of the evaluation of metrics (see Section 4) can be seen in Figures 5 and 6, comparing our method Morphological Superpixel Segmentation (MSS) with SLIC, regular grid watershed segmentation (WS) and also a simple rigid grid segmentation (BOX).

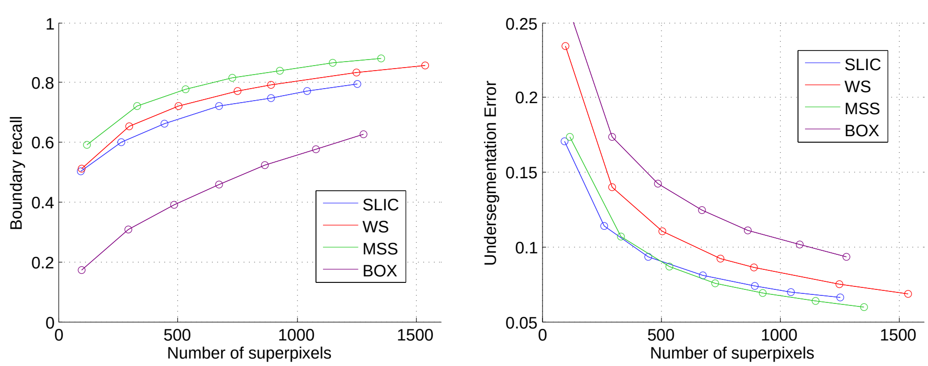

Fig. 5. Left: Boundary recall results. Right: Undersegmentation error.

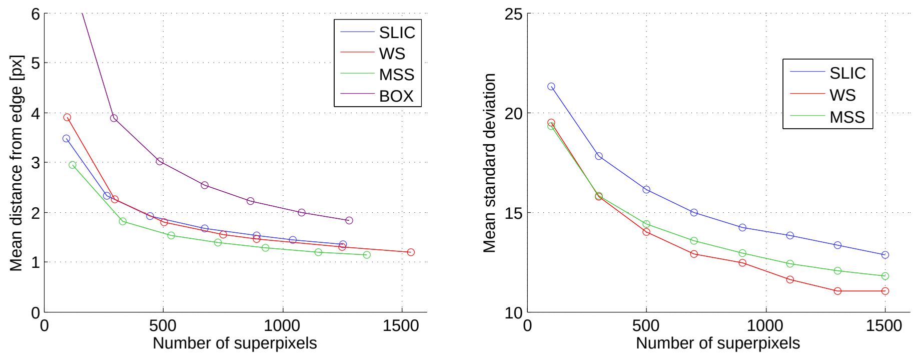

Fig. 6. Left: Mean distance to edge. Right: Intra-cluster variation.

As can be seen from Figure 5, MSS almost consistently achieves better results than the SLIC algorithm in Boundary recall and also in Undersegmentation error. It is also consistently better than the basic WS algorithm.

Our proposed evaluation metrics Mean distance to edge and Intra-cluster variation (Figure 6) are less affected by perpendicularly intersecting boundaries of superpixel which misrepresent the results mainly if the size of superpixel is small. The MDE metric shows that MSS tends to follow human-segmented contours more closely than other evaluated algorithms. The WS algorithm achieves better results in Intra-cluster variation mainly because its markers start as single pixels and always grow in minimal increments, whereas MSS forces an initial selection of regions, which may already contain more variance.

5.2. Timing

The results of the evaluation were obtained on a computer with Intel Core i3 processor clocked at 2.27GHz. The timing results depending on the resolution of the input image can be seen in Table 1. We only compared SLIC, WS and our proposed MSS.

Table 1. The mean time required to compute superpixels depending resolution.

Resolution	SLIC	WS	MSS
192x108	33 ms	4 ms	26 ms
384x216	127 ms	17 ms	33 ms
960x540	766 ms	102 ms	118 ms
1344x756	1511 ms	231 ms	228 ms
1920x1080	3132 ms	462 ms	470 ms

MSS is consistently faster than SLIC, and does not degrade with larger resolutions as fast as SLIC. MSS is significantly slower than WS only on low resolution images, where the morphological image reconstruction causes overhead. However, in higher resolution images the timing tends to level out between MSS and WS because the overhead is compensated by providing a much larger initial marker area, leaving less work to the watershed algorithm.

This favours the proposed method for the use in real time or near real time application of object recognition systems which use superpixels as basic units.

6. Conclusion

We presented a morphological approach to the superpixel segmentation (MSS) which achieves results which are comparable to and better than with the SLIC method.

The evaluation also includes two proposed metrics which aim better to evaluate the superpixel boundary quality in the case of increasing amount of superpixels without a bias of perpendicularly intersecting boundaries of superpixel. Ours proposed superpixel segmentation method is also more suitable for fast object recognition tasks than SLIC, mainly as the resolution of input images increases.

Our future work is towards the improving of the similarity measure between a point and its neighbouring used in watershed algorithm and we will develop an object recognition system based on superpixels.

Acknowledgments

This work was supported by the Grant VEGA 1/0625/14.

References

Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., and Susstrunk, S. (2010). SLIC Superpixels. Technical report, EPFL.

Arbelaez, P., Maire, M., Fowlkes, C., and Malik, J. (2011). Contour detection and hierarchical image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(5):898–916.

Beucher, S. and Lantuejoul, C. (1979). Use of Watersheds in Contour Detection. In *International Workshop on Image Processing: Real-time Edge and Motion Detection/Estimation, Rennes, France.*

Felzenszwalb, P. F. and Huttenlocher, D. P. (2004). Efficient graph-based image segmentation. *Int. J. Comput. Vision*, 59(2):167–181.

Levinshtein, A., Stere, A., Kutulakos, K. N., Fleet, D. J., Dickinson, S. J., and Siddiqi, K. (2009). Turbopixels: Fast superpixels using geometric flows. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(12):2290–2297.

Meyer, F. (1992). Color image segmentation. In *Image Processing and its Applications, 1992., International Conference on*, pages 303–306.

Neubert, P. and Protzel, P. (2012). Superpixel benchmark and comparison. In *Proc. of Forum Bildverarbeitung*. Regensburg, Germany.

Shi, J. and Malik, J. (1997). Normalized cuts and image segmentation. In *Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition (CVPR '97)*, CVPR '97, pages 731–, Washington, DC, USA. IEEE Computer Society.

Veksler, O., Boykov, Y., and Mehrani, P. (2010). Superpixels and supervoxels in an energy optimization framework. In *Proceedings of the 11th European conference on Computer vision: Part V*, ECCV'10, pages 211–224, Berlin, Heidelberg. Springer-Verlag.

Vincent, L. (1993). Morphological grayscale reconstruction in image analysis: applications and efficient algorithms. *Image Processing, IEEE Transactions on*, 2(2):176–201.

Zabih, R. and Kolmogorov, V. (2004). Spatially coherent clustering using graph cuts. In *CVPR (2)*, pages 437–444.